Total Stability in Stable Matching Games

Sushmita Gupta[1], Kazuo Iwama[2], and Shuichi Miyazaki[3]

1 **Institute of Informatics, University of Bergen, Bergen, Norway**
sushmita.gupta@gmail.com
2 **Research Institute for Mathematical Sciences, Kyoto University, Kyoto, Japan**
iwama@kuis.kyoto-u.ac.jp
3 **Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan**
shuichi@media.kyoto-u.ac.jp

Abstract

The stable marriage problem (**SMP**) can be seen as a typical game, where each player wants to obtain the best possible partner by manipulating his/her preference list. Thus the set $\mathbf{Q}$ of preference lists submitted to the matching agency may differ from $\mathbf{P}$, the set of true preference lists. In this paper, we study the stability of the stated lists in $\mathbf{Q}$. If $\mathbf{Q}$ is not Nash equilibrium, i.e., if a player can obtain a strictly better partner (with respect to the preference order in $\mathbf{P}$) by only changing his/her list, then in the view of standard game theory, $\mathbf{Q}$ is vulnerable. In the case of **SMP**, however, we need to consider another factor, namely that all valid matchings should not include any "blocking pairs" with respect to $\mathbf{P}$. Thus, if the above manipulation of a player introduces blocking pairs, it would prevent this manipulation. Consequently, we say $\mathbf{Q}$ is *totally stable* if either $\mathbf{Q}$ is a Nash equilibrium or if any attempt at manipulation by a single player causes blocking pairs with respect to $\mathbf{P}$. We study the complexity of testing the total stability of a stated strategy. It is known that this question is answered in polynomial time if the instance $(\mathbf{P}, \mathbf{Q})$ always satisfies $\mathbf{P} = \mathbf{Q}$. We extend this polynomially solvable class to the general one, where $\mathbf{P}$ and $\mathbf{Q}$ may be arbitrarily different.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases stable matching, Gale-Shapley algorithm, manipulation, stability, Nash equilibrium

Digital Object Identifier 10.4230/LIPIcs.SWAT.2016.23

1 Introduction

Matching under preferences is an extensively studied area of theoretical and empirical research that has a wide range of applications in economics and social sciences. One of the most popular and standard problems in this field is the *stable marriage problem* (**SMP**) introduced by Gale and Shapley [4], where there are two parties; a set of men and a set of women. Each man has a list that orders women according to his preference, and similarly each woman also has a preference ordering for the men. The goal is to find a *stable matching*. A matching is said to be *stable* if it does not have a *blocking pair*, that is, a pair of a man and a woman who prefer each other to their current matching partners. Existence of blocking pairs poses threat to the "stability of the marriage". It is important to note that a matching may be stable in terms of one set of preference lists, but not in terms of another. Hence, whenever there is a possibility of confusion when referring to a stable matching or blocking pairs, we will specify the corresponding preference lists, like "$\mathbf{Q}$-stable" and "$\mathbf{Q}$-blocking pairs" for a set $\mathbf{Q}$ of preference lists. It is well-known that there may be more than one stable matching for a given set of preference lists, where the *men-optimal* and *women-optimal* stable matchings represent the two extremes. These matchings derive their names from the property that, in the men-optimal (women-optimal) stable matching, every man (woman) receives the best possible partner among all the stable matchings. It is well-known that the men-proposing Gale-Shapley algorithm (GS-M, for short) finds the men-optimal stable matching in linear time [4, 6]. Throughout this paper, we focus on only GS-M.

SMP can be seen as a game, in the sense that each player is selfish and wants to obtain the best possible partner, even at the cost of stealing one from the other players. Thus, several game-theoretic issues come into play in this scenario. These issues have been studied mainly from the point of view of economic and market theories (see Roth and Sotomayor [12] for detailed discussions). A fundamental axiom of the selfish game is that players cheat to maximize their individual outcomes. There are a lot of work in this context. Dubins and Freedman [2] showed that, when GS-M is used, no man or a subset of men can misstate their true preference and thereby improve the outcome for all its members. On the other hand, women can manipulate to get better partners when GS-M is used (see [6] for example). Teo *et al.* [13] considered a manipulation by a single woman, and gave an $\mathcal{O}(n^3)$ time algorithm that computes the best matching partner of w who is attainable by applying GS-M to one of the $n!$ permutations of her preference list.

The above approach assumes that the stated preference lists are the same as the true preference lists. As observed above, however, this may be true for men but not for women; a woman w may use a list $Q(w)$ in the game, which is different from her true preference list $P(w)$, in the hope of obtaining a better partner than she obtains when using $P(w)$. This motivates us to consider the case where there are true preferences $\mathbf{P} = (P(M), P(W))$ and stated strategy $\mathbf{Q} = (P(M), Q(W))$, where $P(M)$ and $P(W)$ are the sets of true preference lists of men and women, respectively, and $Q(W)$ is the set of stated preference lists of women. Our aim is to check if $\mathbf{Q}$ is *Nash equilibrium.* A strategy $\mathbf{Q}$ is said to be a Nash equilibrium with respect to strategy $\mathbf{P}$, if no single player can get a strictly better partner (in terms of the true preference list in $\mathbf{P}$) by changing his/her list in $\mathbf{Q}$ while all others use their respective lists in $\mathbf{Q}$.

In this case, we need to consider another factor. Suppose that a woman w, with the true preference $P(w)$ and the stated preference $Q(w)$, may be able to obtain a better partner by using $Q'(w)$ than the one she obtains when $Q(w)$ is used. However, if the resulting matching (which is of course stable with respect to the used preferences) is unstable with respect to $\mathbf{P}$, the matching may be broken and w may lose a partner, so we cannot say this manipulation successful. Therefore, for w's manipulation to be successful, we have to add the condition that the resulting matching is stable with respect to $\mathbf{P}$.

Model. In this article, we study **SMP** in the following game theoretic model. Our game consists of four elements,

(i) a true strategy $\mathbf{P}$,
(ii) an arbitrary stated strategy $\mathbf{Q}$, and
(iii) two different notions of stability for $\mathbf{Q}$:
  (a) one based on the absence of $\mathbf{P}$-blocking pairs, and
  (b) the other based on Nash equilibrium with respect to $\mathbf{P}$.

In this paper, we incorporate all these elements in our model, making it the most general model for stable marriage problems. Throughout this paper, we will *only* deal with strategies that contain the complete and strict preference lists of every man and woman, and so the misstated preferences can only be a permutation of the true preference list.

Men	$P(M)$				Women	$P(W)$				$Q(W)$			
1 :	a	b	c	d	a :	2	3	$\underline{1}$	4	$\underline{3}$	4	1	2
2 :	b	a	c	d	b :	1	$\underline{2}$	3	4	$\underline{1}$	2	3	4
3 :	b	c	a	d	c :	2	$\underline{3}$	4	1	$\underline{2}$	3	4	1
4 :	a	d	b	c	d :	$\underline{4}$	1	2	3	$\underline{4}$	1	2	3

True strategy $\mathbf{P} = (P(M), P(W))$ and stated strategy $\mathbf{Q} = (P(M), Q(W))$.

Figure 1 The men-optimal $\mathbf{Q}$-stable matching is not $\mathbf{P}$-stable, has a blocking pair $(2, a)$.

For a given a strategy $\mathbf{Q} = (P(M), Q(W))$, $\mu_{\mathbf{Q}}$ denotes its men-optimal stable matching. Also, for a player a, we define $\mathscr{S}_{(\mathbf{Q},a)} = \{(Q(-a), Q'(a)) \mid Q'(a)$ is a permutation of $Q(a)\}$ to be the family of strategies obtained by changing the list of a in $\mathbf{Q}$, while all others retain their respective lists in $\mathbf{Q}$. A strategy $\mathbf{Q}$ with respect to the (true) strategy $\mathbf{P}$ is said to be *totally stable* if (a) $\mu_{\mathbf{Q}}$ is $\mathbf{P}$-stable, and (b) for any woman $w \in W$, if $\mathbf{Q'} \in \mathscr{S}_{(\mathbf{Q},w)}$, then either $\mu_{\mathbf{Q'}}$ is not $\mathbf{P}$-stable or $\mu_{\mathbf{Q}}(w) \succeq \mu_{\mathbf{Q'}}(w)$ in $P(w)$, i.e., $\mathbf{Q'}$ yields no better partner for w in terms of $\mathbf{P}$. In plain words, there is no woman who can improve her outcome by changing her list in $\mathbf{Q}$, while the resulting matching is $\mathbf{P}$-stable.

There is a long history of research on the stability of the first kind (discussed while introducing our model), that is matchings that have no blocking pair with respect to a given (fixed) strategy. This includes a beautiful mathematical structure that describes the entire set of stable matchings (see [6], e.g.). However, this knowledge is only applicable when there is only *one* strategy (i.e., $\mathbf{P} = \mathbf{Q}$, in our model). The situation is considerably different when $\mathbf{P} \neq \mathbf{Q}$.

We illustrate the difference between the cases $\mathbf{P} = \mathbf{Q}$ and $\mathbf{P} \neq \mathbf{Q}$ by way of a small example presented in Figure 1. Strategies $\mathbf{P}$ and $\mathbf{Q}$ differ only in the list of a, and the men-optimal $\mathbf{P}$-stable matching is $\mu_{\mathbf{P}} = \{(1, a), (2, b), (3, c), (4, d)\}$. Note that the men-optimal $\mathbf{Q}$-stable matching $\mu_{\mathbf{Q}} = \{(1, b), (2, c), (3, a), (4, d)\}$ is preferred by a to $\mu_{\mathbf{P}}$, in terms of both $\mathbf{P}$ and $\mathbf{Q}$, since $3 \succ 1$ in $P(a)$ and $Q(a)$. By definition, $\mu_{\mathbf{Q}}$ is $\mathbf{Q}$-stable, but as evidenced by the blocking pair $(2, a)$, it is not $\mathbf{P}$-stable.

In comparison, our knowledge about the stability of the second kind, mentioned in our model, is considerably less. As mentioned earlier, it has been known for a long time that when GS-M is used, stating their true preferences is the best strategy for men, but a woman may be able to obtain a better matching partner by manipulating her true list. All these improvements are in terms of the manipulating player's true list. The initial research into the strategic manipulation by women primarily dealt with strategies that were obtained by truncating a player's true list. The possibility of manipulation by permuting the true list (assumed to contain all n players of the opposite kind) has been known for decades (see [6, pg 65]), but to the best of our knowledge, its time complexity was not analysed until Teo *et al.* [13], who gave an $\mathcal{O}(n^3)$ time algorithm to compute the best matching partner of a given woman w. An immediate corollary is that we can test whether the true strategy $\mathbf{P}$ is itself a Nash equilibrium in time $\mathcal{O}(n^4)$. However, [13] does not discuss the stability of the matching obtained by manipulation.

In our case, the situation becomes much more complicated since we have to consider two distinct strategies $\mathbf{P}$ and $\mathbf{Q}$, their respective stabilities, and the interaction between them. We illustrate this scenario by way of a small example in Figure 2. The men-optimal $\mathbf{Q}$-stable matching $\mu_{\mathbf{Q}}$ matches $(1, a)$. The algorithm from [13] applied to a and $\mathbf{Q}$ yields $Q'(a) = [3, 4, 1, 2]$. But, as shown in the earlier example, $(2, a)$ is a $\mathbf{P}$-blocking pair. Thus, seeking efficient algorithms for testing the total stability of a stated strategy seems to be a

Men	$P(M)$				Women	$P(W)$				$Q(W)$			
1 :	a	b	c	d	a :	2	3	$\underline{1}$	4	3	2	$\underline{1}$	4
2 :	b	a	c	d	b :	1	$\underline{2}$	3	4	1	$\underline{2}$	3	4
3 :	b	c	a	d	c :	2	$\underline{3}$	4	1	2	$\underline{3}$	4	1
4 :	a	d	b	c	d :	$\underline{4}$	1	2	3	$\underline{4}$	1	2	3

True strategy $\mathbf{P} = (P(M), P(W))$ and stated strategy $\mathbf{Q} = (P(M), Q(W))$

Figure 2 Algorithm in [13] applied to $\mathbf{Q}$ yields a matching that is not $\mathbf{P}$-stable.

nice algorithmic challenge in the field of matchings under preferences.

By combining [13] and [12, Thm 4.16], we can obtain a polynomial time algorithm to test the total stability when $\mathbf{P} = \mathbf{Q}$. In other words our problem is polynomial time tractable for the special case of $\mathbf{P} = \mathbf{Q}$. The main goal of this paper is to extend this tractability result to the most general setting of $\mathbf{P} \neq \mathbf{Q}$.

1.1 Our Contribution

We show that total stability can be tested in polynomial time. As stated in the Introduction, men do not have any incentive to misstate their true preference ([2] and [12, Theorem 4.10]); consequently, our analysis only considers manipulations by women. Specifically, for two strategies $\mathbf{P} = (P(M), P(W))$ and $\mathbf{Q} = (P(M), Q(W))$ containing complete lists for every man and woman, with $\mathbf{P}$ assumed to be the true strategy, our output should be **No** if there is a woman w and a strategy $\mathbf{Q'} = (Q(-w), Q'(w)) \in \mathscr{S}_{(\mathbf{Q},w)}$ (the family of strategies derived from $\mathbf{Q}$ where only w permutes her preference list) such that (a) $\mu_{\mathbf{Q'}}$ is $\mathbf{P}$-stable, and (b) w obtains a strictly better partner in $\mu_{\mathbf{Q'}}$, in terms of $P(w)$ (notationally expressed as $\mu_{\mathbf{Q'}}(w) \succ \mu_{\mathbf{Q}}(w)$, in $P(w)$). If there is no such strategy $\mathbf{Q'} \in \mathscr{S}_{(\mathbf{Q},w)}$ for any woman w, then the answer should be **Yes**.

The obvious brute-force method is to check all the $n!$ permutations as the list $Q'(w)$; once $Q'(w)$ is fixed, computing $\mu_{\mathbf{Q'}}$ and checking if it is $\mathbf{P}$-stable can be done in $\mathcal{O}(n^2)$ time. For a polynomial time algorithm, we need to do the following two computations without examining all the $n!$ permutations for $Q'(w)$: (i) obtaining all possible (by changing only her list) partners of w who are better (in terms of P) than w's partner in $\mu_{\mathbf{Q}}$, and (ii) for a man m found in (i), we need to search for a permutation $Q'(w)$ such that w is matched to m and $\mu_{\mathbf{Q'}}$ is $\mathbf{P}$-stable.

We wish to point out that the result in [13] is not enough for the first task, since it only gives the best partner in terms of $\mathbf{Q}$, and the matching outcome may not be $\mathbf{P}$-stable in general, as depicted by Figure 1. Note that if a is satisfied with the second best partner, 2, then she could use a list $[2, 3, 4, 1]$ and the resulting matching, $(a, 2), (b, 1), (c, 3), (d, 4)$, is now $\mathbf{P}$-stable. Thus it does not suffice to merely detect the best possible partner of a. Also, note that potential partners of women other than w may be arbitrary and there can be (exponentially) many lists for w that result in matching w to m, some of which may yield a $\mathbf{P}$-stable matching, while others may not, even if m is fixed as a target.

The basic idea of our algorithm is as follows: For task (ii), we obtain a result (Theorem 5) which proves that if two permutations $Q'_1(w)$ and $Q'_2(w)$ are available as $Q'(w)$ (both matching w to m), then $\mu_{\mathbf{Q'_1}} = \mu_{\mathbf{Q'_2}}$. Hence, if any one list gives a $\mathbf{P}$-stable outcome, then so do the others. Thus, for algorithmic purposes, it is enough to consider *any* arbitrary single permutation as $Q'(w)$. We believe that this result is interesting in its own right, and could be useful in other scenarios. For task (i) we can use an algorithm based on the same idea as [13], but its correctness proof is quite different from the original one, that heavily relies on the fact that they are only interested in the best manipulated partner of w. For our purpose, we need to detect all attainable partners.

1.2 Related Work

In addition to [13] that we have discussed in details, there are other relevant works that we can point to.

For a stated strategy $\mathbf{Q}$, Dubins and Freedman [2] proved that there is no coalition C of men who have a manipulation strategy $\mathbf{P'} = (P(-C), P'(C))$, so that the outcome is $\mathbf{P'}$-stable and is strictly better than $\mu_{\mathbf{P}}$ in terms of $\mathbf{P}$, for each $m \in C$. Demange *et al.* [1] extend this result to include women in the coalition C, showing that there is no $\mathbf{P'}$-stable matching μ' such that every player $a \in C$ prefers $\mu'(a)$ (in terms of $P(a)$) to his/her partner in *every* $\mathbf{P}$-stable matching.

For truncation strategies, it was shown by Gale and Sotomayor [5] that if there are at least two $\mathbf{P}$-stable matchings, then there is a woman w who has a unilateral manipulation strategy $\mathbf{Q'} \in \mathscr{S}_{(\mathbf{Q},w)}$ that gives a strictly better outcome than $\mu_{\mathbf{P}}$. If $C = W$, then there is a truncation strategy $\mathbf{P'} = (P(M), P'(W))$ such that $\mu_{\mathbf{P'}}$ is the women-optimal $\mathbf{P}$-stable matching. Considerable work on truncation strategies have been undertaken (see [3, 11] for motivations and applications). In fact, up until the late 1980s, analyses of manipulation strategies of women centred almost exclusively around truncation strategies.

Immorlica and Mahidian [7] show that with high probability, truthfulness is the best strategy for any individual player, assuming everybody else is being truthful as well. In their model, the men's preference lists may have ties but the lengths are bounded by a constant, and are drawn from an arbitrary probability distribution, while the women's lists are arbitrary and complete.

Kobayashi and Matsui [8] consider the possibility that a coalition C of women have a manipulation strategy $\mathbf{P'} = (P(M), P'(W))$ containing complete lists, such that $\mu_{\mathbf{P'}}$ yields specific partnerships for the members of C. The situation manifests in two specific forms, depending on the nature of the input. In the first case, the input consists of the complete lists of all men, a partial matching (some agents may be unmatched) μ', and complete lists of the subset of women who are unmatched in μ', denoted by $W \setminus C$. The problem is to test whether there exists a permutation strategy for each woman in C, such that for the combined strategy $\mathbf{P'} = (P(-C), P'(C))$, $\mu_{\mathbf{P'}}$ is a perfect matching that extends μ'. In the second case, the input consists of the lists of all men, a perfect matching μ, and lists for women in $W \setminus C$. The problem is to test if there are permutation strategies for the women in C such that strategy $\mathbf{P'} = (P(-C), P'(C))$ yields $\mu'_{\mathbf{P}} = \mu$. They present polynomial-time algorithms for both problems.

Pini *et al.* [9] show that for an arbitrary instance of **SMP**, there is a stable matching mechanism for which it is NP-hard to find a manipulation strategy.

Roth [10] had shown that if a strategy $\mathbf{Q} = (P(M), Q(W))$ is a Nash equilibrium with respect to $\mathbf{P}$, then the matching $\mu_{\mathbf{Q}}$ is $\mathbf{P}$-stable. The proof discussed in [12] allows women to truncate their preference lists as a means of manipulation. This result holds even if all players are restricted to using complete lists in their true, stated and manipulated strategies. We use this result without a proof, as it is similar to the second approach described in [12, pg 101].

2 Preliminaries

We will always use M to denote the set of n men $\{m_1, m_2, \ldots, m_n\}$ and W the set of n women $\{w_1, w_2, \ldots, w_n\}$. Our matching mechanism is the men-proposing Gale-Shapley algorithm

(GS-M in short), which proceeds as follows: On the men's side, a man who is not yet matched to a woman, *proposes* to the woman who is at the top of his current list, which is obtained by removing all the women who have already rejected him. On the woman's side, when a woman w receives a proposal from a man m, she *accepts* the proposal if it is her first proposal, or if she prefers m to her current partner m'. If w prefers her current partner m' to m, then w *rejects* m. If m is rejected by w, then m will start issuing proposals, and this process continues until there is no man left who is unmatched. For more details, see [6]. At any stage of GS-M, if there are two or more unmatched men, then we set the convention that in this group the man with the smallest index is the first to propose. This removes the possibility of arbitrary tie-breaking, and thus, makes the algorithm purely deterministic.

A *strategy* $\mathbf{Q}$ is a set of *preference lists* (or simply *lists*) of all the men in M and all the women in W. For a person x in $M \cup W$, $Q(x)$ denotes the x's list in the strategy $\mathbf{Q}$. For a given strategy $\mathbf{Q}$, suppose that *only* w changes her list from $Q(w)$ to $Q'(w)$. We denote the resulting strategy by $\mathbf{Q'} = (Q(-w), Q'(w))$, and use $\mathscr{S}_{(\mathbf{Q},w)}$ to denote the family of all such strategies $\mathbf{Q'}$. Note that all lists considered in this article are complete, i.e., they are permutations of n men or n women.

Let $\mathbf{Q}$ be a strategy. If w prefers m_1 to m_2 in $Q(w)$, then we write "$m_1 \succ m_2$ in $Q(w)$". We use $m_1 \succeq m_2$ if $m_1 \succ m_2$ or $m_1 = m_2$. Let μ be a (perfect) matching between M and W. Then $\mu(p)$ denotes the partner of a person p. A pair (m, w) of a man and a woman is called a $\mathbf{Q}$-*blocking pair* if $w \succ \mu(m)$ in $Q(m)$ and $m \succ \mu(w)$ in $Q(w)$. We say that μ is $\mathbf{Q}$-*stable* if there is no $\mathbf{Q}$-blocking pair.

For a strategy $\mathbf{Q}$, $\mu_{\mathbf{Q}}$ denotes the man-optimal stable matching, computed by the Gale-Shapley algorithm. If a man m proposes to a woman w during this procedure, then we say that m is *active* in $Q(w)$ (formally speaking we should say m is active in $Q(w)$ during the computation of $\mu_{\mathbf{Q}}$, but for the sake of brevity, we will omit strategy $\mathbf{Q}$ when it is obvious from the context.)

Recall that a woman w changes her list $Q(w)$ for the purpose of manipulation. For a subset $M' \subseteq M$, let I be an ordering of men in M'. Then, $Q(I; w)$ denotes a permutation of $Q(w)$, where the men in M' are at the front in the order in which they appear in I. An ordered (sub)list, such as I, is called a *tuple*, and for any given tuple I, we define $Q(I; w) = [I, Q(w) \backslash I]$. For example, if $Q(w) = [1, 2, 3, 4, 5, 6]$ and $I = [5, 2]$, then $Q(I; w) = [5, 2, 1, 3, 4, 6]$. Now we are ready to introduce our main concept.

We are given a strategy $\mathbf{Q}$ and a *true* strategy $\mathbf{P}$. Then for a woman $w \in W$, a strategy $\mathbf{Q'} \in \mathscr{Q}_w$ is said to be a *unilateral manipulation strategy* of w, if $\mu_{\mathbf{Q'}}(w) \succ \mu_{\mathbf{Q}}(w)$ in $P(w)$, i.e., w strictly prefers the outcome of $\mu_{\mathbf{Q'}}$ to $\mu_{\mathbf{Q}}$ with respect to her true preference/strategy. If, furthermore, $\mu_{\mathbf{Q'}}$ is a $\mathbf{P}$-stable matching, then $\mathbf{Q'}$ is said to be a $\mathbf{P}$-*stable manipulation strategy* of w. A strategy $\mathbf{Q}$ is said to be totally stable if there does not exist a $w \in W$ who has a $\mathbf{P}$-stable manipulation strategy $(Q(-w), Q'(w)) \in \mathscr{Q}_w$. In this paper, we consider the following problem.

Problem: TOTAL STABILITY
Input: True strategy $\mathbf{P} = (P(M), P(W))$ and stated strategy $\mathbf{Q} = (P(M), Q(W))$
Question: Is $\mathbf{Q}$ totally stable?

3 Listing active men

Now our goal is to design an algorithm that, for two given strategies, a stated strategy $\mathbf{Q}$ and a true strategy $\mathbf{P}$, answers if $\mathbf{Q}$ is totally stable. To do so, we first design an algorithm that

Algorithm 1: $\mathcal{A}(\mathbf{Q}, w)$ [13]

Input: Strategy $\mathbf{Q} = (P(M), Q(W))$, and a woman $w \in W$

Output: Sets $N_w(\mathbf{Q}) = \{m \in M \mid \exists\ \mathbf{Q'} \in \mathscr{S}_{(\mathbf{Q},w)}$ that yields $\mu_{\mathbf{Q'}}(w) = m\}$, and
$L_w(\mathbf{Q}) = \{Q'(m;w) \mid m \in N_w(\mathbf{Q}),\ \mathbf{Q'} = (Q(-w), Q'(m;w))$ yields $\mu_{\mathbf{Q'}}(w) = m\}$

Let x_1 be the first active man in $Q(w)$
Let $N \leftarrow \{x_1\}$ and $L \leftarrow \{Q(x_1;w)\}$
Explore$(Q(x_1;w))$
return (N, L)

Procedure **Explore**$(Q'(x, I; w))$

Let $A \leftarrow$ {men who are active in $Q'(x, I; w)$ after x}
foreach $y \in A \setminus N$ **do**
    $N \leftarrow N \cup \{y\}$ and $L \leftarrow L \cup \{Q'(y, x, I; w)\}$
    Explore$(Q'(y, x, I; w))$

outputs the set $N_w(\mathbf{Q})$ of all possible partners m of a given fixed woman w such that there is a (manipulated) strategy $\mathbf{Q'} = (Q(-w), Q'(w))$, for which the men-optimal stable matching will match w to m. By using this algorithm n times, we can obtain $N_{w_1}(\mathbf{Q}), \ldots, N_{w_n}(\mathbf{Q})$. The use of this set to prove our main result is explained in the next section.

Let us consider Algorithm 1, which is basically the same as the one given by Teo *et al.* [13]: Suppose that $Q(w) = [1, 2, 3, 4, 5, 6, 7, 8]$ and the first proposal comes from man 5. Then the algorithm adds 5 to N and calls **Explore**$(Q(5;w))$: it executes the men-proposing Gale-Shapley algorithm (GS-M, in short) after moving 5 to the front of the list $Q(w)$. In general, procedure **Explore** takes as a parameter $Q(x, I; w)$, a preference list of w. As per our notation, x is at the front of this list, followed by the sublist I and then the rest of the men, thus, defining the strategy $\mathbf{Q'} = (Q(-w), Q(x, I; w))$. **Explore**$(Q(x, I; w))$ executes GS-M for the strategy $\mathbf{Q'}$ and produces the set of men A who propose to w after x. Now for each $y \in A$, we check if y is "new" (i.e., not yet in N). If so, we add y to N and call **Explore** recursively after moving y to the top of $Q(x, I; w)$; else, we do nothing.

Since **Explore** is called only once for each man in N, its time complexity is obviously at most $n \times T(\text{GS})$, where $T(\text{GS})$ is the time complexity of one execution of GS-M, thus, overall it is $\mathcal{O}(n^3)$. The nontrivial part is the correctness of the argument, which we shall prove now.

▶ **Theorem 1.** *For a strategy* $\mathbf{Q}$ *and a woman* $w \in W$*, Algorithm 1 produces* $N = \{m \in M \mid \exists \mathbf{Q'} \in \mathcal{Q}_w,\ s.t.\ \mu_{\mathbf{Q'}}(w) = m\}$ *and for each* $m \in N$*, a list* $Q(m, I; w)$ *such that, for some partial list* I*,* m *is active in* $Q(I; w)$*.*

Proof. Let $Q'(w)$ be an arbitrary permutation of n men and $\mathbf{Q'}$ the strategy $(Q(-w), Q'(w))$. It is enough to prove if a man $x \in M$ proposes to w during the computation of $\mu_{\mathbf{Q'}}$ (i.e., x is active in $Q'(w)$), then x is added to N during the execution of Algorithm 1.

Here we need two new definitions: Suppose that $x_1, x_2, \ldots, x_t$ is a sequence of men who proposed to w (in this order) during the computation of $\mu_{\mathbf{Q'}}$. Then this sequence is called an *active sequence* for $Q'(w)$, denoted by $\text{AS}'(w)$. Also define $y_1, y_2, \ldots, y_s$ as a maximal subsequence of $\text{AS}'(w)$ such that $y_1 = x_1$ and for $i \geq 2$, y_i is the first element after y_{i-1} such that $y_i \succ y_{i-1}$ in $Q'(w)$. This is called the *increasing active subsequence* for $Q'(w)$ and is denoted by $\text{IAS}'(w)$. As an example, let $Q'(w) = [1, 2, 3, 4, 5, 6, 7, 8, 9]$

and $\mathrm{AS}'(w) = 5, 6, 3, 4, 2, 8$. Then $\mathrm{IAS}'(w) = 5, 3, 2$. Now consider a different list $Q''(w) = [1, 2, 3, 5, 9, 8, 4, 6, 7]$, thus, $Q'(w) \neq Q''(w)$. However, we can observe that the active sequence and the increasing active subsequence for $Q''(w)$ are identical to those of $Q'(w)$, for the following reasons. The lists in $\mathbf{Q}'$ and $\mathbf{Q}''$ are the same except that of w's, so the first proposal for w must come from the same man regardless of w's list. Since the man 5 is accepted by w in both executions, the next proposal should also be from the same man 6. Now since $5 \succ 6$ in both $Q'(w)$ and $Q''(w)$, 6 is rejected in both $Q'(w)$ and $Q''(w)$ and thus, the next proposal must also be same, and so on. This observation leads us to the following lemma.

▶ **Lemma 2.** *For strategies $\mathbf{Q}', \mathbf{Q}'' \in \mathcal{Q}_w$, let $x_1, x_2, \ldots, x_p$ and $u_1, u_2, \ldots, u_q$ denote the active sequences for $Q'(w)$ and $Q''(w)$ respectively, and let $y_1, y_2, \ldots, y_s$ and $v_1, v_2, \ldots, v_t$ denote the corresponding increasing active subsequence. Then, the following conditions must hold.*

(a) $x_1 = y_1 = u_1 = v_1$.

(b) *For an arbitrary l ($l \leq p$ and $l \leq q$), we consider the prefixes of the active sequences up to position l and the prefixes of the corresponding increasing active subsequences, denoted by $y_1, ..., y_j$ and $v_1, ..., v_j$. Then, if $x_i = u_i$ for all $i \leq l$ and $y_k = v_k$ for all $k \leq j$, then $x_{l+1} = u_{l+1}$.*

Proof. By definition, $x_1 = y_1$ and $u_1 = v_1$. Recall that all the lists in Q' and Q'' are the same except those for w. Furthermore, we use a fixed tie-breaking protocol in the deterministic GS algorithm. Hence, $x_1 = u_1$ follows directly.

To prove condition (b), let $y_2 = x_{i_1+1}, y_3 = x_{i_2+1}, \ldots$, and so on. Then we can write $\mathrm{AS}'(w)$ as follows, where $x_2, \ldots, x_{i_1}, x_{i_1+2}, \ldots, x_{i_2}, \ldots$ may be empty.

$$\mathrm{AS}'(w) = y_1, x_2, \ldots, x_{i_1}, y_2, x_{i_1+2}, \ldots, x_{i_2}, \ldots, y_j, x_{i_{j-1}+2}, \ldots, x_l, x_{l+1}, \ldots$$

Now one can see that y_1 is accepted, all of $x_2 \ldots, x_{i_1}$ are rejected since they are after y_1 in the list by definition. This continues as y_2 is accepted, $x_{i_1+2}, \ldots, x_{i_2}$ rejected, and so on. Now, consider $\mathrm{AS}''(w)$, depicted below.

$$\mathrm{AS}''(w) = v_1, u_2, \ldots, u_{i_1}, v_2, u_{i_1+2}, \ldots, u_{i_2}, \ldots, v_j, u_{i_{j-1}+2}, \ldots, u_l, u_{l+1}, \ldots$$

By the assumption, these two sequences are identical up to position l, so acceptance or rejection for each proposal follows identically, as discussed above. Therefore, the *configuration* (see below) of GS-M for $\mathbf{Q}'$ at the moment when x_l proposes to w and the configuration for $\mathbf{Q}''$ when u_l proposes to w are exactly the same. A configuration consists of (i) the lists of all men (recall that some entries are removed when proposals are rejected), (ii) the set of single men, and (iii) the current temporal matching partner of each woman. (Formally this should be shown by induction, but it is straightforward and is omitted). Also the acceptance/rejection for x_l and u_l is the same. Thus in either case, the execution of the (deterministic) GS-M is exactly the same for $\mathbf{Q}'$ and $\mathbf{Q}''$ until w receives proposal from x_{l+1} and u_{l+1}, respectively. Hence, x_{l+1} and u_{l+1}, should be equal and the lemma is proved. ◀

Now let us look at the execution sequence of Algorithm 1 while comparing it with the execution sequence of GS-M on $\mathbf{Q}'$. Let the active sequence and increasing active sequence for $\mathbf{Q}'$ be $\mathrm{AS}'(w) = x_1, x_2, \ldots, x_p$ and $\mathrm{IAS}'(w) = y_1, y_2, \ldots, y_s$, respectively. By Lemma 2, the first proposal to w is always y_1, so the algorithm starts with **Explore**$(Q(y_1; w))$ (we simply say the algorithm invokes $Q(y_1; w)$), and $N = \{y_1\}$, at the very beginning.

Now we note that it is quite easy to see that the active sequence for $Q(y_1;w)$ should be $y_1, x_2, \ldots, x_{i_1}, \ldots$, i.e., it should be identical to that of $Q'(w)$ up to the position i_1, with i_1 defined as in the proof of Lemma 2. The reason is as follows. We already know the first active man is always y_1 and that is also the first symbol in the increasing active sequence of both. Thus we can use Lemma 2 to conclude that the second symbol should also be the same, since x_2 is not in $\text{IAS}'(w)$, meaning that it is rejected, which is also the same in $Q(y_1;w)$ since y_1 is at the top of the sequence. Thus the third symbol is the same in both and so on up to position i_1. Then the next symbol in $\text{AS}'(w)$ is y_2 and it is also active in $Q(y_1;w)$, meaning $Q(y_2,y_1;w)$ is invoked by the algorithm. (The algorithm also invokes $Q(x_2,y_1;w)$, $Q(x_3,y_1;w), \ldots, Q(x_{i_1},y_1;w)$, but these are not important for us at this moment.)

We again consider the active sequence for $Q(y_2,y_1;w)$ and by the same argument presented earlier, we can conclude that it is identical to $\text{AS}'(w)$ up to position i_2 and so y_3 is found to be an active man. Hence, $Q(y_3,y_2,y_1;w)$ is invoked if y_3 was not already present in N. Continuing like this, we note that if $Q(y_s,y_{s-1},\ldots,y_1;w)$ is invoked, then we are done since its active sequence is identical to that of $Q'(w)$. However, this case happens only if each y_i $(2 \le i \le s)$, is a brand new active man found during the invocation of $Q(y_{i-1},\ldots,y_1;w)$. If one of them is not new then the subsequent lists are not invoked, and yet, we are assured due to Lemma 3 that Algorithm 1 will detect all the active men in $Q'(w)$.

Lemma 3 is rather surprising and may be of independent interest. For two lists $Q'(w)$ and $Q''(w)$, that are distinct and arbitrary orderings on men, we assume nothing about the execution of GS-M on the two lists except that a particular man x is active in both lists. Yet, we are able to show that a man who proposes to w when $Q'(x;w)$ is used must also propose when $Q''(x;w)$ is used. This result allows us to focus solely on active men that have been discovered in the current invocation of **Explore**, thereby restricting the number of recursion steps to $\mathcal{O}(n)$.

▶ **Lemma 3.** *For two distinct strategies $\mathbf{Q}'$ and $\mathbf{Q}''$ in $\mathscr{S}_{(\mathbf{Q},w)}$, suppose that x is active in both $Q'(w)$ and $Q''(w)$. Then a man who is active in $Q'(x;w)$ is also active in $Q''(x;w)$.*

Proof. Consider the strategies $\mathbf{Q_1} = (Q'(-w), Q'(x;w))$ and $\mathbf{Q_2} = (Q''(-w), Q''(x;w))$. Let y denote a man who is active in $Q'(x;w)$ but not in $Q''(x;w)$. Then, $\mu_{\mathbf{Q_2}}(y) \succ w \succ \mu_{\mathbf{Q_1}}(y)$ in $P(y)$.

Note that $\mu_{\mathbf{Q_1}}(w) = x$ and $\mu_{\mathbf{Q_2}}(w) = x$. Clearly, any $\mathbf{Q_1}$-blocking pair in $\mu_{\mathbf{Q_2}}$ must involve w, as otherwise it would also be a $\mathbf{Q}_2$-blocking pair. However, since x is at the top of w's list in $\mathbf{Q}_1$, w cannot be in a $\mathbf{Q_1}$-blocking pair, implying that $\mu_{\mathbf{Q_2}}$ is a $\mathbf{Q}_1$-stable matching. Since $\mu_{\mathbf{Q_1}}$ is the men-optimal stable matching for $\mathbf{Q}_1$, we have that $\mu_{\mathbf{Q_1}}(y) \succeq \mu_{\mathbf{Q_2}}(y)$ in $P(y)$. This contradicts the fact we have shown earlier, and hence y must be active in $Q''(x;w)$. ◀

The next lemma completes the proof. We give one more notation, where y_is are men defined in Lemma 2.

$$Q_1(w) = Q_1(y_1;w), \text{ and } Q_{j+1}(w) = [y_{j+1}, Q_j(w) \setminus \{y_j\}], \text{ for } 1 \le j \le s-1.$$

▶ **Lemma 4.** *For each j $(1 \le j \le s)$, the algorithm invokes $Q(y_j, I; w)$ for some tuple I, and each man in $\left\{x_{i_{j-1}+2}, \ldots, x_{i_j}, y_{j+1}\right\}$ is active in it.*

Proof. We prove this result by induction on y_i. The base case has been already proved, since for y_1, it has been shown earlier that $Q(y_1;w)$ is invoked at the beginning and every man in $\{x_2, \ldots, x_{i_1}, y_2\}$ is active in $Q(y_1;w)$ after y_1.

Suppose that the induction hypothesis holds for y_t, where $t \le s-2$, i.e., for some tuple I, $Q(y_t, I; w)$ is invoked, and each man in $\left\{x_{i_{t-1}+2}, \ldots, x_{i_t}, y_{t+1}\right\}$ is active in it. We will complete the proof by showing that the hypothesis holds for y_{t+1}. If y_{t+1} is "new", i.e., it is added to N during the invocation of $Q(y_t, I; w)$, then $Q(y_{t+1}, y_t, I; w)$ is invoked subsequently. Using the fact that y_{t+1} is active in both $Q_{t+1}(w)$ and $Q(y_t, I; w)$, and all the men in $\{x_{i_t+2}, \dots, y_{t+2}\}$ are active in $Q_{t+1}(w)$ after y_{t+1}, Lemma 3 applied to each of them implies that they are also active in $Q(y_{t+1}, y_t, I; w)$. Hence, for this case, the hypothesis is proved for y_{t+1}.

If y_{t+1} is already in N when $Q(y_t, I; w)$ is invoked, then for some tuple I', y_{t+1} should have been added to N during the invocation of $Q(I'; w)$. Thus, $Q(y_{t+1}, I'; w)$ would have been invoked prior to $Q(y_t, I; w)$. Using the same argument (on $Q_{t+1}(w)$ and $Q(y_{t+1}, I'; w)$) that we used for the earlier case, we conclude that even for this case, the hypothesis holds for y_{t+1}. ◀

Thus, we have shown that all the men in $\mathrm{AS}'(w)$ are active somewhere during the execution of the algorithm and thus, all are present in N at the end of the execution. This completes the proof of Theorem 1. ◀

4 Algorithm to test if a strategy is totally stable

In this section, we consider the problem of deciding, for a given true strategy $\mathbf{P}$ and a stated strategy $\mathbf{Q}$, whether $\mathbf{Q}$ is totally stable. We show that this problem is solvable in time $\mathcal{O}(n^4)$. Algorithm 2 uses Algorithm 1 as a subroutine.

Algorithm 2: Algorithm for TOTAL STABILITY.

Input: True strategy $\mathbf{P}$, stated strategy $\mathbf{Q}$, and the set of women W.
Output: Answers "Yes", if $\mathbf{Q}$ is totally stable, else "No".

foreach $w \in W$ **do**
    Run Algorithm 1 on input $(\mathbf{Q}, w)$ to obtain $N_w(\mathbf{Q})$ and $L_w(\mathbf{Q})$
    Let $\widetilde{N} \leftarrow \{m \in N_w(\mathbf{Q})$ s.t. w prefers m to $\mu_{\mathbf{Q}}(w)$, in $P(w)\}$
    foreach $m \in \widetilde{N}$ **do**
        Let $Q(m, I; w) \in L_w(\mathbf{Q})$ be the list that yields (m, w) as a matched pair
        Let μ be the men-optimal $(Q(-w), Q(m, I; w))$-stable matching
        if *μ is $\mathbf{P}$-stable* **then**
            return *"No"*
return *"Yes"*

The following result is of independent interest as it implies that Algorithm 1 on the input $(\mathbf{Q}, w)$ generates *all matchings* that can be attained by changing w's preference list. In particular, it answers if for any given $w \in W$, there exists a strategy $\mathbf{Q}' \in \mathscr{S}_{(\mathbf{Q},w)}$ such that $\mu_{\mathbf{Q}'}(w) \succ \mu_{\mathbf{Q}}(w)$ in $P(w)$, and the matching $\mu_{\mathbf{Q}'}$ is $\mathbf{P}$-stable matching. As a consequence, using Algorithm 1 as a subroutine, Algorithm 2 is able to solve TOTAL STABILITY.

▶ **Theorem 5.** *For any (fixed) man $\overline{m}$, if there exists a permutation $Q'(w)$ of a woman w's list $Q(w)$ such that the strategy $\mathbf{Q}' = (Q(-w), Q'(w))$ yields a matching that matches w to $\overline{m}$, then that matching is unique.*

Proof. Suppose that $Q'(w)$ (defined in the theorem) is an arbitrary strategy of w to obtain $\overline{m}$, and let μ' denote the men-optimal $\mathbf{Q}'$-stable matching. Algorithm 1 applied to input $(\mathbf{Q}, w)$

computes a list $Q(\overline{m}, I; w)$ such that w attains $\overline{m}$ by the strategy $\mathbf{Q}^* = (Q(-w), Q(\overline{m}, I; w))$. Note that $\overline{m}$ appears at the front of the list $Q^*(w) = Q(\overline{m}, I; w)$. Let μ^* be the men-optimal $\mathbf{Q}^*$-stable matching. Our goal is to show that $\mu' = \mu^*$.

▶ **Claim 6.** *For each m, $\mu^*(m) \succeq \mu'(m)$ in $Q'(m)$.*

Proof. We begin by showing that μ' is $\mathbf{Q}^*$-stable. We know that μ' is $\mathbf{Q}'$-stable, and $\mathbf{Q}'$ and $\mathbf{Q}^*$ differ only in w's list. Hence, if there is a $\mathbf{Q}^*$-blocking pair in μ', then it must contain w. However, this is impossible since w is matched with $\overline{m}$, who is at the front of the list $Q^*(w)$. Therefore, μ' must be $\mathbf{Q}^*$-stable

Since μ^* is the men-optimal $\mathbf{Q}^*$-stable matching and μ' is a $\mathbf{Q}^*$-stable matching, we have that, for each man m, $\mu^*(m) \succeq \mu'(m)$ in $Q^*(m)$. Since $Q^*(m) = Q'(m)$ for each man m, the claim is proved. ◀

▶ **Claim 7.** *For each m, $\mu'(m) \succeq \mu^*(m)$ in $Q'(m)$.*

Proof. As for Claim 6, we begin by showing that μ^* is $\mathbf{Q}'$-stable. Suppose that it is not. Then there is a $\mathbf{Q}'$-blocking pair in μ^*, and it includes w for the same reason as in the proof of Claim 6. Let (m', w) denote a $\mathbf{Q}'$-blocking pair in μ^*. Then, $w \succ \mu^*(m')$ in $Q'(m')$, and $m' \succ \mu^*(w)$ in $Q'(w)$.

Using Claim 6, we have $w \succ \mu'(m')$ in $Q'(m')$, and $\mu^*(w) = \mu'(w) = \overline{m}$ by definition. Hence, (m', w) is a $\mathbf{Q}'$-blocking pair in μ', a contradiction. Again, for the same reason as in the proof of Claim 6, we can conclude that $\mu'(m) \succeq \mu^*(m)$ in $Q'(m)$ for each man m. ◀

By Claims 6 and 7, we have $\mu'(m) = \mu^*(m)$ for each man m. Thus, $\mu' = \mu^*$, completing the proof of Theorem 5. ◀

▶ **Theorem 8.** *Algorithm 2 solves* Total Stability *in $\mathcal{O}(n^4)$ time.*

Proof. Suppose that Algorithm 2 outputs "No". Then, it implies that a $\mathbf{P}$-stable manipulation strategy was found, and therefore $\mathbf{Q}$ is not totally stable. For the opposite direction, suppose that $\mathbf{Q}$ is not totally stable and there exists a woman w who has a manipulation strategy $\mathbf{Q}'$ such that $\mu_{\mathbf{Q}'}$ is $\mathbf{P}$-stable. Then, man $\mu_{\mathbf{Q}'}(w)$ is added to $\widetilde{N}$ when Algorithm 1 is executed on the input $(\mathbf{Q}, w)$. By Theorem 5 the matching $\mu_{\mathbf{Q}'}$ is uniquely defined, i.e., there is a unique matching resulting from a manipulation strategy of w that results in the matched pair $(w, \mu_{\mathbf{Q}'}(w))$. Since $\mu_{\mathbf{Q}'}$ is $\mathbf{P}$-stable, Algorithm 2 will output "No." This proves the correctness of Algorithm 2.

We claim that the time complexity of Algorithm 1 is $\mathcal{O}(n^3)$. This is because the size of the set N is at most n, and is computed iteratively by executing GS-M once for each man $m \in N$. Since the running time of GS-M is $\mathcal{O}(n^2)$, the running time of Algorithm 1 is $\mathcal{O}(n^3)$. Algorithm 2 executes Algorithm 1 for each woman $w \in W$. Hence, the running time of Algorithm 2 is $\mathcal{O}(n^4)$. ◀

5 Conclusions

We leave the question of manipulation by a group of women as an avenue of further research. In particular, we would like to answer in polynomial time (1) if a stated strategy $\mathbf{Q}$ is totally stable against manipulations by a subset of women, and (2) if a given subset of women $W' \subseteq W$ have a manipulation strategy that yields a $\mathbf{P}$-stable matching and gives each of them a better partner than the one given by the stated strategy. In this article, we solved both of these problems for the special case of manipulation by one woman acting on her own.

Acknowledgments. The authors would like to thank the anonymous reviewers for their valuable comments. This work was supported by JSPS KAKENHI Grant Numbers 24500013 and 25240002.

References

1 G. Demange, D. Gale, and M. Sotomayor. A further remark on the stable matching problem. *Discrete Applied Mathematics*, 16:217–222, 1987.

2 L. E. Dubins and D. A. Freedman. Machiavelli and the Gale-Shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494, 1981.

3 L. Ehlers. Truncation strategies in matching markets. *Mathematics of Operations Research*, 33(2):327–335, 2008.

4 D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.

5 D. Gale and M. Sotomayor. Ms. Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 92(4):261–268, 1985.

6 D. Gusfield and R. W. Irving. *The Stable Marriage Problem-Structure and Algorithm.* The MIT Press, 1989.

7 N. Immorlica and M. Mahidian. Marriage, honesty and stability. In *Proceedings of SODA'05*, pages 53–62, 2005.

8 H. Kobayashi and T. Matsui. Cheating strategies for the Gale-Shapley algorithm with complete preference lists. *Algorithmica*, 58:151–169, 2010.

9 M. S. Piny, F. Rossi, K. B. Veneble, and T. Walsh. Manipulation complexity and gender neutrality in in stable marriage procedures. *Auton. Agent Multi-Agent Systems*, 22(183–199), 2011.

10 A. E. Roth. Misrepresentation and stability in the marriage problem. *Journal of Economic Theory*, 34:383–387, 1984.

11 A. E. Roth and U. G. Rothblum. Truncation strategies in matching markets-in search of advice for participants. *Econometrica*, 67(1):21–43, 1999.

12 A. E. Roth and M. Sotomayor. *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis.* Cambridge Univ. Press, 1990.

13 C-P. Teo, J. Sethuraman, and W-P. Tan. Gale-Shapley stable marriage problem revisited: Strategic issues and applications. *Management Science*, 47(9):1252–1267, 2001.